TRANSPARENT
PROCESS

EMBO
*reports*

# U2AF$^{65}$ assemblies drive sequence-specific splice site recognition

Manel Tari[1], Valérie Manceau[2,†], Jean de Matha Salone[1], Asaki Kobayashi[1], David Pastré[1] & Alexandre Maucuer[1,*] (ID)

## Abstract

The essential splicing factor U2AF$^{65}$ is known to help anchoring U2 snRNP at the branch site. Its C-terminal UHM domain interacts with ULM motifs of SF3b155, an U2 snRNP protein. Here, we report a cooperative binding of U2AF$^{65}$ and the related protein CAPERα to the multi-ULM domain of SF3b155. In addition, we show that the RS domain of U2AF$^{65}$ drives a liquid–liquid phase separation that is amplified by intronic RNA with repeated pyrimidine tracts. In cells, knockdown of either U2AF$^{65}$ or CAPERα improves the inclusion of cassette exons that are preceded by such repeated pyrimidine-rich motifs. These results support a model in which liquid-like assemblies of U2AF$^{65}$ and CAPERα on repetitive pyrimidine-rich RNA sequences are driven by their RS domains, and facilitate the recruitment of the multi-ULM domain of SF3b155. We anticipate that posttranslational modifications and proteins recruited in dynamical U2AF$^{65}$ and CAPERα condensates may further contribute to the complex mechanisms leading to specific splice site choice that occurs in cells.

## Introduction

Splicing of intervening sequences from eukaryotic pre-mRNAs is achieved by a macromolecular machinery called the spliceosome that assembles in a stepwise manner (reviewed in Refs [1,2]). In an early step of assembly, the U2AF small subunit (U2AF$^{35}$) binds the conserved AG dinucleotide just preceding the 3′ splice site [3–5] (Fig 1A). The U2AF large subunit (U2AF$^{65}$) binds the polypyrimidine tract (PPT) through its tandem RNA recognition motifs (RRM) [3,6–8]. U2AF$^{65}$ also possesses a C-terminal atypical RRM motif called "U2AF Homology Motif" (UHM) [9,10] (Fig 1B). This atypical RRM has lost its RNA-binding ability, but

it interacts with an N-terminal "UHM Ligand Motif" (ULM) of SF1 [11–13].

In the second step of spliceosome assembly, SF1 that was bound to the branchpoint sequence is replaced by the U2 snRNA-containing ribonucleoprotein (U2 snRNP) with the help of U2AF$^{65}$. The N-terminal arginine- and serine-rich (RS) domain of U2AF$^{65}$ contacts the branchpoint sequence (BPS) and favors the formation of a U2 snRNA-BPS duplex [14]. The U2AF$^{65}$ UHM domain engages interactions with ULM motifs of the U2 snRNP subunit SF3b155 [15]. UHM domains have inherited from RRM domains their overall structural pattern with two helices packed against a beta sheet (Fig 1C). They present however important differences enabling specific interactions [10]. In particular, the space between the helices forms a hydrophobic binding pocket for a tryptophan residue of the ULM in the protein partner. UHM domains are also found in a subset of proteins, mainly splicing factors (here called U2AF homology motif splicing factors or UHMSFs) including the transcription and splicing regulators PUF60 and CAPERα as well as the splicing and DNA repair factor SPF45 [10].

A single ULM domain is present in SF1 that binds the U2AF$^{65}$ UHM during the pioneer branchpoint region recognition [11–13,16–19] but, intriguingly, seven ULM copies were identified in SF3b155 (Fig 1B). Five of these potential ULMs have been shown to bind U2AF$^{65}$ *in vitro* [20–22] (Appendix Fig S1). PUF60, CAPERα, and SPF45 were also reported to bind the SF3b155 multi-ULM domain (review in Ref. [10]). In addition, structures of CAPERα and SPF45 UHMs in complex with the ULM corresponding to tryptophan 338 have been determined [23,24]. The possibility that multiple copies of an UHMSF are simultaneously bound to SF3b155 is suggested by the fact that UHMSF binding to ULMs present in SF3b155 does not present a strict specificity (Appendix Fig S1). For example, the differences in affinity of CAPERα or U2AF$^{65}$ for binding their two preferred ULMs in SF3b155 are small. Altogether, the binding of combinations of UHMSFs to the multiple ULMs in SF3b155 is likely to be functionally important, but has not been deeply explored so far [10].

Among UHMSFs, CAPERα, also called RBM39, has the highest structural similarity with U2AF$^{65}$ [25] and is the subject of a particular attention due to its role in cancer-related splicing events and in tumor progression [26,27]. The regulation of specific splicing events

---

1 SABNP, Univ Evry, INSERM U1204, Université Paris-Saclay, Evry, France
2 Institut Necker Enfants Malades (INEM), Inserm U1151 − CNRS UMR 8253, Université Paris Descartes, Paris, France
*Corresponding author. Tel: +33 169870187; E-mail: alexandre.maucuer@inserm.fr
†Present address: Faculty of Medicine, Institut Necker Enfants Malades (INEM), Inserm U1151−CNRS UMR 8253, University Paris Descartes, Sorbonne Paris Cité, Paris, France
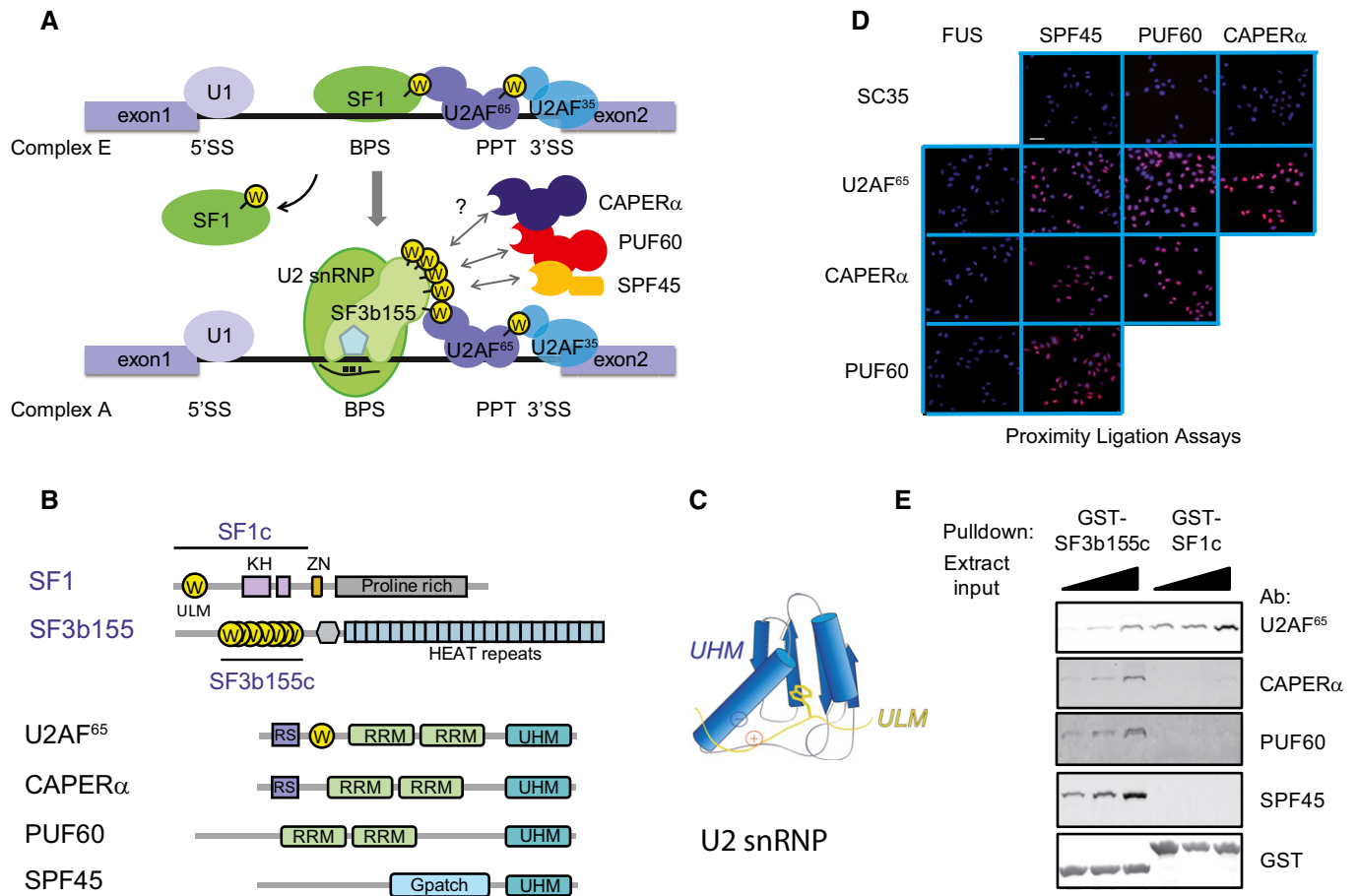
**Figure 1. Interactions of UHM domain proteins in early spliceosome assembly.**

A   Schematic of early splicing factors recruitment to the splice sites. SS, splice site; BPS, branchpoint sequence; PPT, polypyrimidine tract; W, tryptophan. The ULM–UHM contact between SF1 and U2AF[65] in early complex E is replaced by a similar SF3b155–U2AF[65] interaction in complex A when the U2 snRNA becomes base-paired with the branchpoint.

B   Domain organization of 3′ splice site splicing factors in this study. UHM, U2AF homology motif; ULM, UHM ligand motif; KH, hnRNPK homology motif; ZN, zinc finger; RS, arginine- and serine-rich motif; RRM, RNA recognition motif. The regions used for SF3b155 and SF1 in this study are depicted (SF3b155c: human SF3b155, residues 190–344, and SF1c: human SF1, residues 1–255).

C   Schematic of the UHM–ULM interaction structure. A conserved tryptophan accommodated into a hydrophobic pocket and ionic interactions with acidic residues of helix A are main features of this interaction.

D   Proximity ligation assays between splicing factors in HeLa cells. Different pairs of antibodies against the indicated splicing factors were used with concentrations adjusted in immunofluorescence experiments (Appendix Fig S2). Blue, DAPI staining; red, PLA signal (scale bar: 10 μm).

E   GST pulldown using recombinant GST-SF3b155c or GST-SF1c and increasing amounts of RNase-treated extract from HEK293 cells. Precipitated material was analyzed by Western blotting (representative results of two experiments).

by CAPERα [25,27], which may be altered in acute myeloid leukemia, constitutes a putative therapeutic route [28]. In a previous work, we illustrated the possibility that multiple CAPERα molecules bind SF3b155 in a cooperative manner [23]. This further supported the hypothesis that different combinations of UHMSFs cooperatively bind SF3b155 as part of a modular mechanism for 3′ splice site recognition.

Here, we interrogate whether homotypic or heterotypic UHMSF associations to the multi-ULM domain of SF3b155 could play a functional role in splicing. We reveal a particular relationship of U2AF[65] and CAPERα that cooperatively interact with the multi-ULM domain of SF3b155 in a manner that depends on the RS domain of U2AF[65]. We show that U2AF[65] and CAPERα self-aggregate in different conditions *in vitro*, as known for other RS domain-containing proteins

[29–31]. Using phase-contrast microscopy, we further demonstrate that the RS domain of U2AF[65] can drive a liquid–liquid phase separation (LLPS) to form liquid droplets stabilized by the presence of repeated pyrimidine tracts. LLPS enables the formation of dynamical compartments in which the exchange of constituents with the surrounding is possible [32]. Such compartments have been already identified in the nucleus of mammalian cells as, for example, nucleoli, paraspeckles, or Cajal bodies (review in Ref. [33]). The association of U2AF[65] and CAPERα in dynamical liquid-like assemblies may thus reveal an unknown mechanism of splicing regulation. To document this point, we analyzed splicing events upon knockdown of four UHMSFs. The results indicate a functional relationship of U2AF[65] and CAPERα. Notably, the knockdown of either U2AF[65] or CAPERα improves the inclusion of cassette exons that are preceded

by repeated pyrimidine tracts. At the light of these *in vitro* and in cells results, we propose a mechanistic model in which the recruitment of U2 snRNP at the 3′ intronic sequences is regulated by liquid-like assemblies of U2AF[65] and CAPERα generated by self-attracting RS domains, multiple UHM–ULM interactions with SF3B155, and bindings of RRMs to repeated pyrimidine-rich sequences.

# Results

### Cooperative binding of CAPERα and U2AF[65] to the SF3b155 multi-ULM domain

Conventional immunolabeling shows expression and nuclear localization of U2AF[65], CAPERα, PUF60, and SPF45 in HeLa cells (Appendix Fig S2). We determined antibodies concentrations yielding similar fluorescence intensities for each protein, and used such concentrations to perform proximity ligation assays. We observed that the colocalization of these four UHMSFs occurs in cells. In contrast, no colocalization signal was detected for these UHMSFs with FUS or SC35, a nuclear RNA-binding protein, and a splicing factor, respectively, used as controls (Fig 1D). These data extend a previous FRET analysis that demonstrated the proximity of overexpressed U2AF[65] and CAPERα [34], and support the possibility that pairs of UHMSFs have coordinated actions in splicing. A recent report has now established that the protein tat-SF1 can also form a UHM–ULM interaction structure with SF3b155, which might extend the complexity of the network of UHMSF interactions *in vivo* [35].

*In vitro*, we then compared the binding of U2AF[65], CAPERα, PUF60, and SPF45 to SF3b155. Pulldown assays using RNase-treated cell extracts showed that these four UHMSFs could bind to the ULM-containing domain of SF3b155 fused to glutathione S-transferase (GST-SF3b155c, Fig 1E). To probe a putative cooperative binding of the UHMSFs with SF3b155, a similar pulldown assay was performed in the presence of increasing amounts of recombinant U2AF[65]. U2AF[65] reinforced the recruitment of CAPERα to GST-SF3b155c but not that of PUF60 and SPF45 (Fig 2A). A similar effect was observed when adding a high-affinity complex of U2AF[65] with a minimal domain of U2AF[35] that should not allow the access of CAPERα to the ULM of U2AF[65] (Fig 2A, right panel and Appendix Fig S3) [36,37]. U2AF[65] addition did not lead to an efficient pulldown of CAPERα with GST-SF1c (GST-SF1_1-255; Fig EV1A). This comparison between GST-SF1c and GST-SF3b155c suggested that the multiplicity of ULMs that is specific to SF3b155 is responsible for a cooperative recruitment of CAPERα and U2AF[65] on SF3b155c. To further test this requirement of multiple ULMs for cooperativity, we produced GST-SF3b155c with different combinations of ULM tryptophan mutations. W338, W200, and W293 are the best binders for U2AF[65] in decreasing order (Fig 2B and Appendix Fig S1). We first mutated each of the seven tryptophan for alanine and then reintroduced successively W338, W200, and W293 (Fig 2C). Only GST-SF3B155c with three ULMs supported a dramatic increase in recruitment of CAPERα upon addition of U2AF[65] similar to what is observed with wild-type SF3b155 (Fig 2D).

To further document the role of multiple ULMs in binding U2AF[65] and CAPERα, we compared their interaction with SF3b155c mutants with only one of the three best U2AF[65] ligands (W200, W293, and W338), the three combinations of two of these

tryptophan residues, and finally the combination of these three tryptophan residues. We observed that the combination of three ULMs is far more efficient in recruiting U2AF[65] (Fig 2E; e.g., binding to W200-293-338 was twice as much as the sum of binding to W200 and to W293-338). Similar results were observed for CAPERα. Altogether, the increased binding of CAPERα to SF3b155c in the presence of U2AF[65] and the nonlinearity of U2AF[65] and CAPERα binding as a function of ULM number can be explained by the cooperative binding of CAPERα and U2AF[65] to multiple ULMs of SF3b155.

To identify which domain of U2AF[65] is responsible for this cooperativity, we analyzed the binding of CAPERα to SF3b155c in the presence of U2AF[65] with different N-terminal truncations. The N-terminal region of U2AF[65], containing its RS domain, appeared critical to favor CAPERα interaction with SF3b155c (Fig EV1B). We therefore expressed an RS domain-deleted form of U2AF[65] in HEK293 cells and tested its binding to SF3b155c. Removal of the RS domain strongly impaired the association of U2AF[65] with SF3b155c (Fig 2F). Altogether, these pulldown experiments suggested that the RS domain of U2AF[65] is necessary to stabilize CAPERα and U2AF[65] assemblies on the multi-ULM domain of SF3b155 (Fig 2G).

### Liquid–liquid phase separation driven by the U2AF[65] RS domain

RS domains are low complexity domains that are known to drive aggregation of proteins, probably by generating multivalent interactions among large assemblies [29]. In agreement, we observed, in sedimentation assays, the common properties of U2AF[65] and CAPERα, which both present a RS domain in their N-terminal region (Fig 1B), to form large assemblies on their own (Fig 3A). We then asked whether U2AF[65] and CAPERα assemblies could be considered as localized liquid droplets formed at the 3′ splice site. In agreement with this hypothesis, phase-contrast microscopy revealed the presence of droplets of U2AF[65] which were sensitive to salt, SDS, concentration, and PEG-induced crowding as expected for a LLPS mechanism (Fig 3B–E). This LLPS mechanism was also supported by the dynamics of these objects as evidenced by time-lapse microscopy (Fig 3B bottom panel). Such a LLPS generally originates from multivalent interactions supported by low complexity domains. To ascertain the role of the low complexity RS domain in LLPS of U2AF[65], we purified this RS domain fused to GST. The results indicate that liquid droplets are formed with the RS domain alone (Fig 3B, right panel). In addition, deletion of the RS domain resulted in reduced sedimentation of U2AF[65] (Fig 3A), indicating that this domain is mostly responsible for the formation of U2AF[65] assemblies. Together with our pulldown results, these observations of LLPS support a function for U2AF[65] and CAPERα assemblies, and our hypothesis that their RS domains engage multivalent interactions that can stabilize complexes formed by U2AF[65] and CAPERα on the multi-ULM surface of SF3b155.

### Functional relationship of CAPERα and U2AF[65] in cassette exon splicing

We hypothesized that assemblies of CAPERα and U2AF[65] could mediate a particular recognition of intronic regions with repeated binding sites for their RNA-binding domains. Indeed while a motif of nine nucleotides is sufficient to accommodate the tandem RRMs of U2AF[65] [38], much longer pyrimidine-rich sequences can be
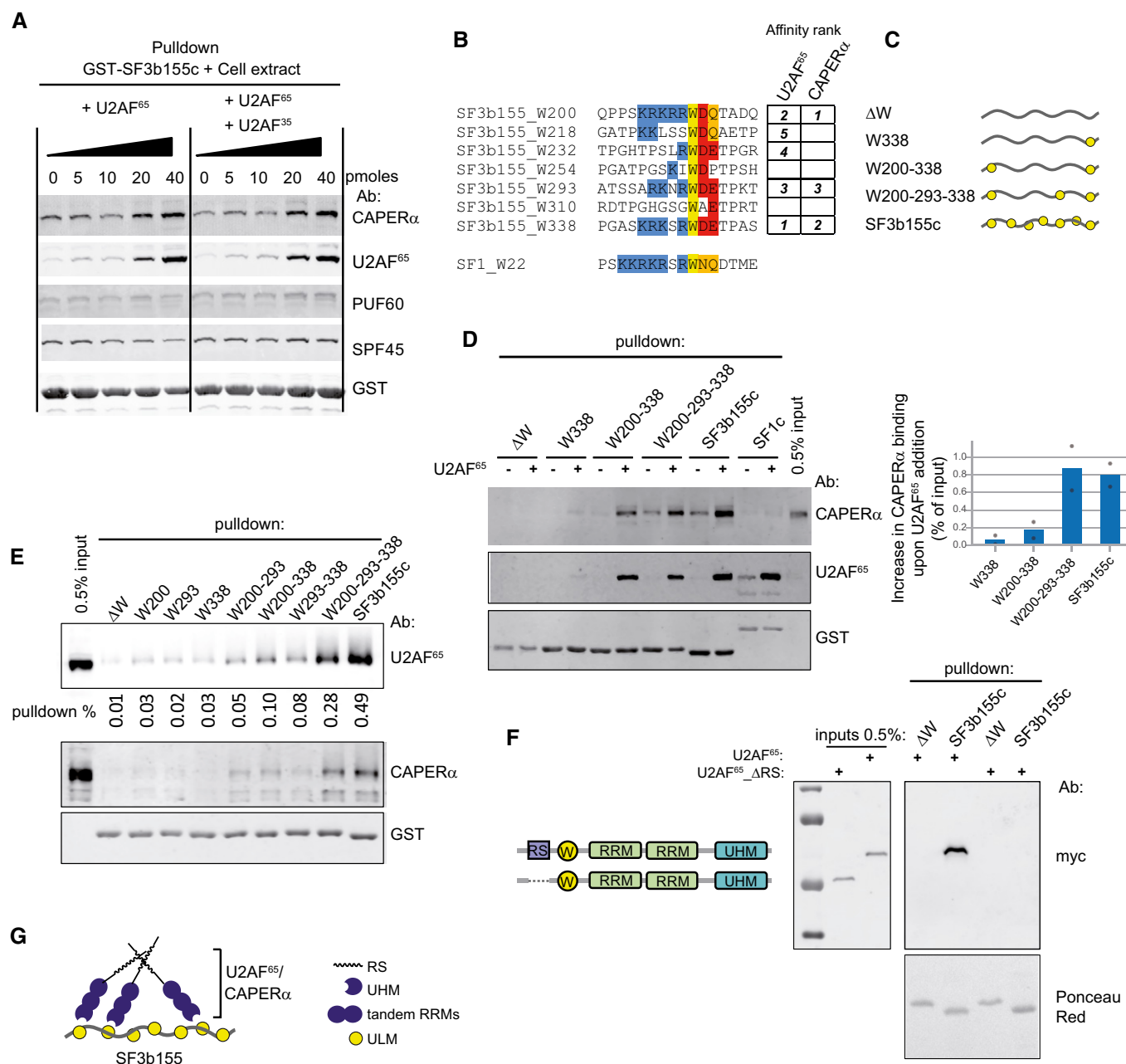
**Figure 2. Cooperative interactions of UHM domain proteins on the SF3b155 multi-ULM domain.**

A  GST pulldown using 25 pmoles of recombinant GST-SF3b155c (see Fig 1B for protein domains) with 200 μg of protein cell extract and addition of increasing amounts of recombinant U2AF65 or a stoichiometric mixture of U2AF65 and a minimal U2AF65-binding domain of U2AF35 (human U2AF35 residues 43–146, stoichiometry analyzed in Appendix Fig S3) [36]. Precipitated material was analyzed for UHMSF content by Western blotting (n = 3).

B  Alignment of the known and putative ULMs in the N-terminal region of SF3b155 with that of SF1. The evaluated rank of each ULM considering its affinity for U2AF65 or CAPERα is indicated on the right (see Appendix Fig S1). The typical ULMs present basic residues (highlighted in blue) N-terminal to the tryptophan, and acidic (red) or glutamine/asparagine residues (orange) C-terminal to the tryptophan.

C  Schematics of representative mutated forms of SF3b155c used for panel (D), with 1, 2, or 3 tryptophan residues in the multi-ULM domain of SF3b155.

D  Pulldown assays as in (A) with different GST-fusion proteins and addition or not of 40 pmoles of U2AF65. The mutated forms with 0, 1, 2, or 3 tryptophan are compared with wild-type SF3b155c and SF1c. Quantification of CAPERα pulldown in two independent experiments shows that SF3b155c mutant with three but not with two tryptophan supports the strong cooperative effect of adding U2AF65.

E  GST-pulldown assays using 40 pmoles of recombinant GST-SF3b155c or the mutant forms indicated, with 200 μg of protein cell extract. The fraction of precipitated U2AF65 was measured while the detection of CAPERα on the same blot was noisier. Blotting with the anti-GST antibody was performed to ascertain that similar quantities of the baits were recovered and for normalization. Two experiments gave similar results.

F  Pulldown assay as in (E) with extract of HEK293 cells expressing myc-U2AF65 or a mutated form lacking the RS domain (myc-U2AF65_ΔRS). The interacting proteins were detected by immunoblotting with anti-myc antibody. Similar recovery of GST-SF3b155c and the mutated form lacking tryptophan (ΔW) was checked by Ponceau red staining of the membrane (complete gel shown in Appendix Fig S4; representative results of two independent experiments).

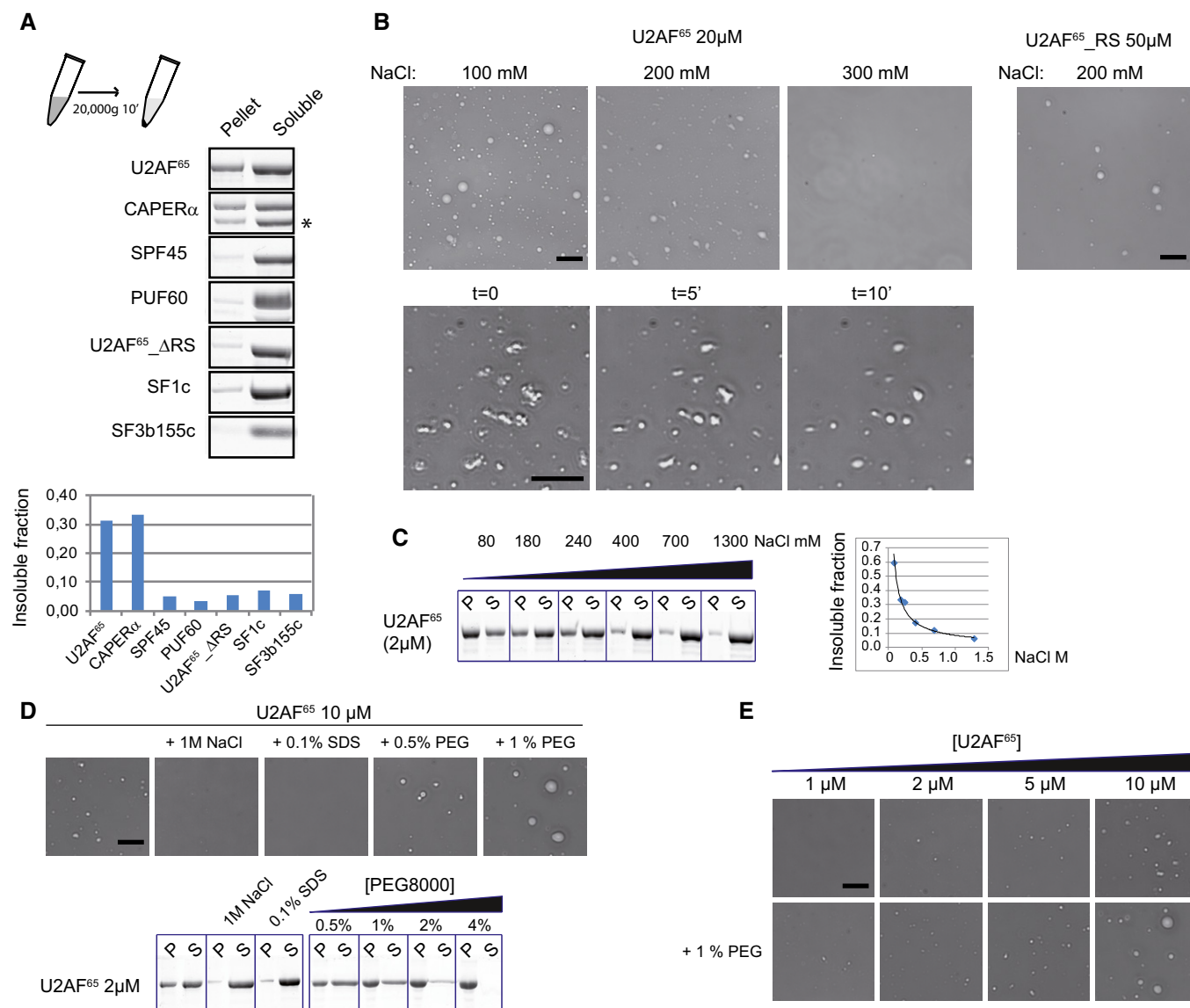G  Model for cooperative interaction of CAPERα and U2AF65 with the SF3b155 multi-ULM domain.

**Figure 3. Large molecular assemblies of U2AF65.**

A Sedimentation assay for the indicated purified proteins at 2 μM in 120 mM NaCl and 20 mM Hepes pH 7.5. After sedimentation, proteins were analyzed by SDS–PAGE and Coomassie staining (See Appendix Fig S5 for full gel). An asterisk indicates a degradation product for CAPERα (P, pellet; S, supernatant; U2AF65_ΔRS: human U2AF65 lacking residues 23–63).

B Phase-contrast observation of U2AF65 and RS domain aggregation (RS domain: human U2AF65 residues 27–62). Increasing salt concentrations reduces droplets size and number (upper panels). Loading under the coverslip eventually leads to shearing of U2AF65 droplets that recover their round shape over time (video, lower panels). Scale bars: 10 μm throughout the figure.

C Effect of salt on U2AF65 condensation in a sedimentation assay as in (A).

D Effect of salt, sodium dodecyl sulfate (SDS), and polyethylene glycol on U2AF65 LLPS and sedimentation showing correlated responses in the two assays (in the presence of 20 mM Hepes pH 7.5 and 140 mM or 1 M NaCl when indicated). Scale bar: 10 μm.

E U2AF65 LLPS at different concentrations without or in the presence of 1% polyethylene glycol to induce molecular crowding (with 20 mM Hepes pH 7.5 and 140 mM NaCl). Scale bar: 10 μm.

found in 5′ flanking regions (region directly 5′ to the exon). First, using a bioinformatics pipeline, we selected 5,629 robust cassette exons from the annotated human exons in the NCBI Reference Sequence (RefSeq) library [39]. Then, we scored the 5′ flanking regions of these exons for number and composition of small pyrimidine stretches (SPY) representing potential binding sites for the U2AF65 and CAPERα tandem RRMs [28,38,40] (See Materials and Methods for detailed definition of this SPY score). Sixty-five exons with low SPY score or high SPY score were finally selected (Fig 4A and Appendix Fig S6). To reduce the expression levels of each UHMSF in HeLa cells, we selected two efficient shRNAs per UHMSF (Appendix Figs S7 and S8 and Appendix Table S1). Then, the inclusion of each of the cassette exons of our collection was measured in each knockdown condition using microfluidic-based high-density

qPCR (Fig 4 and Appendix Fig S9). For each exon and each knockdown, we scored the "Splicing Index" and a "Splicing Change" value as the relative effect of the knockdown to the maximal possible increase in inclusion (Fig 4B). Analyses of Splicing Changes demonstrated a globally positive effect of U2AF[65], CAPERα, and PUF60 on exon inclusion but an inhibitory action of SPF45 in agreement with previous functional characterization of these factors [24,41,42] (Fig 4C). U2AF[65] and CAPERα presented a statistically significant overlap among decreased inclusion events upon knockdown (Fig 4D). We then compared the Splicing Changes for all the cassette exons in response to the different knockdowns (Fig 4E). Spearman's correlation analyses of these data confirmed the particular relationship between U2AF[65] and CAPERα consistently with recent siRNA-mediated knockdown and RNA-Seq experiments [42].

## U2AF[65] and CAPERα differentially regulate cassette exons with repeated pyrimidine tracts

We tested whether the impact of U2AF[65] and CAPERα knockdown on cassette exon inclusion was dependent on the abundance of putative binding sites for U2AF[65] in the 5′ flanking region (SPY score). Alternative splicing of a cassette exon implies that splicing occurs after transcription of the downstream exon in a significant number of occasions. In these situations, the splice sites for the cassette exon and that of the downstream exon are competing for splicing. Therefore, the difference between the SPY score for the cassette exon and the next downstream exon could determine the inclusion rate of a cassette exon (ΔSPY score, see Fig 4A and Materials and Methods). We found a significant positive correlation of "Splicing Changes" upon knockdown of CAPERα and U2AF[65] with the ΔSPY scores (Fig 4F, Spearman rho = 0.40, $P$ = 1.6E-03 and rho = 0.33, $P$ = 8.8E-03, respectively). To generalize these observations, we performed similar analyses using available RNA-Seq data for K562 [43], HeLa [44], MCF7 [42], and HEK293T [45] cells. We aligned reads to our library of sequences corresponding to 5,629 cassette exon inclusions and exclusions. These analyses confirmed different effects of the knockdowns (Fig EV2A), the particular correlation of CAPERα and U2AF[65] (Fig EV2B), as well as the positive correlation of U2AF[65] and CAPERα knockdown with the ΔSPY score (Figs 4G and EV2C). To further support the role of the SPY-rich intronic sequences in favoring the recognition of cassette exon when U2AF[65] or CAPERα levels were reduced, we prepared minigenes corresponding to two cassette exons with strong differences in ΔSPY scores and with opposite responses to CAPERα or U2AF[65] knockdown (low ΔSPY score:TJP1 exon 20; high ΔSPY score: DNM2 exon 14, see Fig 4B). We also switched the 5′ flanking regions in two additional minigenes (Appendix Fig S10). Replacing the TJP1 SPY-poor 5′ flanking region with the DNM2 SPY-rich sequence modifies the response of the TJP1 exon so that it behaves like the DNM2 exon, i.e., its inclusion increases upon U2AF[65] or CAPERα depletion. Symmetrically, replacing the 5′ flanking region of DNM2 by that of TJP1 leads to reduced inclusion of the cassette exon upon U2AF[65] or CAPERα knockdown. Therefore, reducing the pools of U2AF[65] or CAPERα favors the recognition of a repetitive pyrimidine-rich 3′ intronic sequence over a pyrimidine-poor sequence by the splicing machinery. This observation can be explained by cooperative mechanisms associated with recognition of multiple pyrimidine tracts by U2AF[65] and CAPERα assemblies (see Discussion).

## Interactions of large assemblies of U2AF[65] and CAPERα with long pyrimidine-rich RNAs

To test our hypothesis that recognition of SPY-rich sequences might be related to the formation of U2AF[65] and CAPERα assemblies, we prepared 130 base RNAs corresponding to the SPY-poor 5′ flanking intronic region of TJP1 exon 20 and the SPY-rich 5′ flanking region of DNM2 exon 14 (Fig 5A and Appendix Fig S11). Mobility shift assay could not reveal major differences in the binding of U2AF[65] to these RNAs (Appendix Figs S12 and S13). In contrast, the addition of the SPY-rich DNM2 RNA but not that of the SPY-poor TJP1 RNA increased dramatically the formation of U2AF[65] assemblies (Fig 5B and C). Furthermore, this SPY-rich RNA partly cosedimented with U2AF[65] assemblies (Fig 5C, compare lanes 3 and 4 with lanes 5 and 6). Analyses of four additional RNAs confirmed the correlation between SPY content and the ability of an RNA to exacerbate U2AF[65] sedimentation (Appendix Fig S14). Insertion in the TJP1 sequence, of established target sequences for U2AF[65] such as the AdML [38] or SXL sequence [46], was not mimicking the DNM2 RNA in increasing U2AF[65] sedimentation (Appendix Fig S15). The affinity of DNM2 RNA for U2AF[65] is therefore apparently not the primary determinant for its specific effect on U2AF[65] sedimentation. In addition, in minigene experiments, insertion of these preferred binding sequences for U2AF[65] in the TJP1 intron leads to efficient inclusion in basal conditions, and not to an increased inclusion upon U2AF[65] or CAPERα depletion that was only observed with the repeated pyrimidine-rich DNM2 sequence (Appendix Fig S10D).

The increased formation of U2AF[65] assemblies in the presence of the DNM2 SPY-rich intronic RNA was not observed with U2AF[65] lacking its RS domain or with the RS domain alone (Appendix Fig S16). Therefore, RNA increases RS-dependent sedimentation most probably by favoring local concentration of U2AF[65] on repeated binding sites for its RRMs. Adding an excess of DNM2 RNA could also reduce the content of insoluble U2AF[65], in agreement with an expected negative effect on phase separation of lowering local U2AF[65] concentrations through dispersion on many RNA molecules (Appendix Fig S17). This increase in U2AF[65] solubility in the presence of high RNA concentrations could also partly result from interaction of RNA with the RS domain, as proposed for different SR proteins [29]. Finally, addition of RNase in the assay demonstrated that the sedimentation of U2AF[65] without added RNA was not due to bacterial RNA remaining associated with U2AF[65] during purification (Appendix Fig S18).

## SF3b155 multi-ULM domain further promotes assemblies of U2AF[65] in the presence of pyrimidine-rich RNA

Having determined the concentration-dependent positive effect of SPY-rich RNAs on the formation of U2AF[65] assemblies, we tested the impact of SF3b155 in this process. The formation of U2AF[65] assemblies upon DNM2 RNA addition was exacerbated by the presence of the SF3b155 multi-ULM domain but not by a mutated domain lacking the essential tryptophan residues of the seven ULM motifs (tryptophan residues replaced by alanine; SF3b155c_ΔW; Fig 5C lanes 11–14). While SF3b155c was mainly soluble even at high concentration, it was efficiently recruited to U2AF[65] assemblies in the presence of SPY-rich but not SPY-poor RNA (Fig 5C, compare lanes 9 and 11). SF3b155c in turn enhanced the preferential
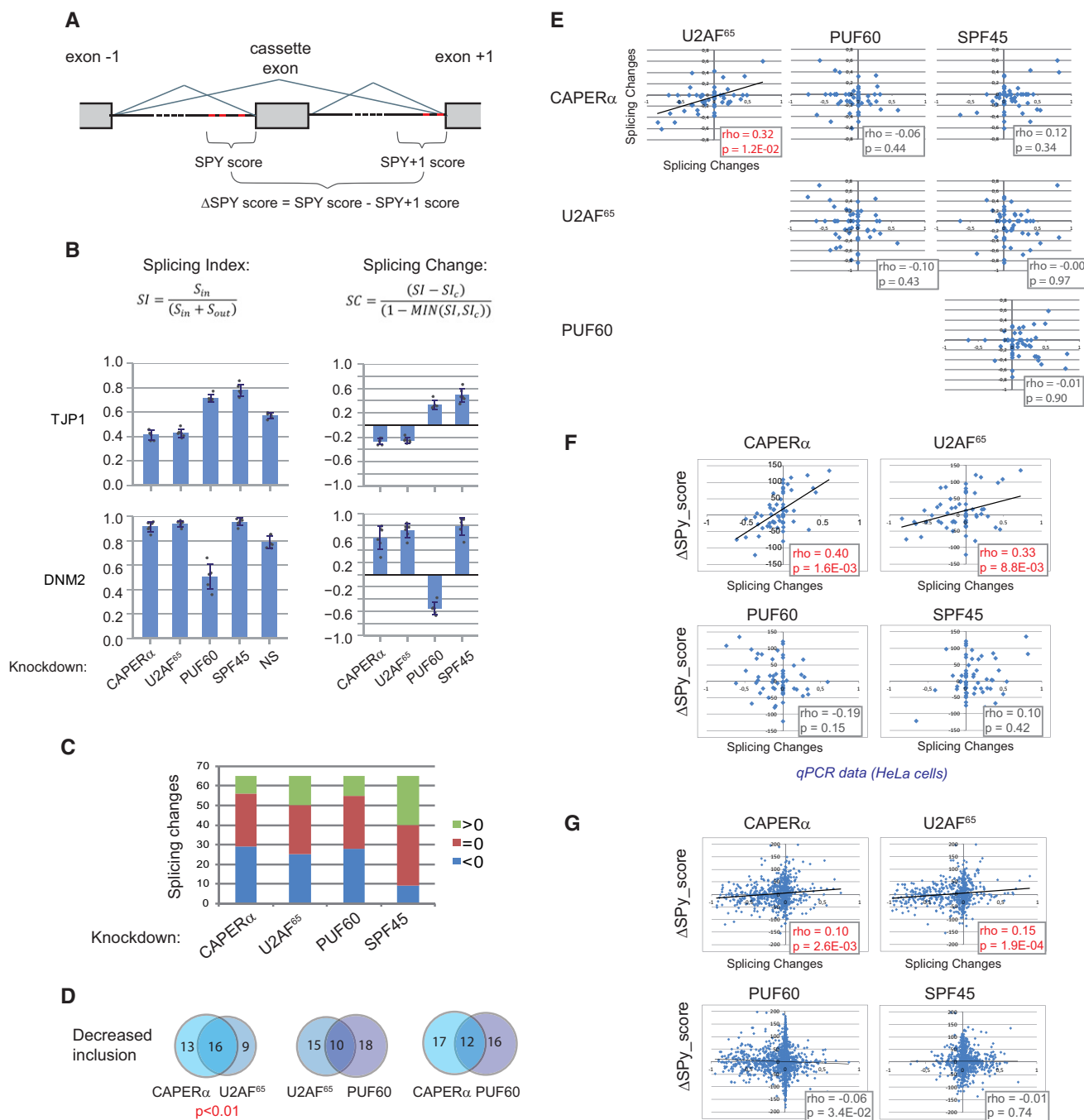
**Figure 4. Alternative splicing of cassette exons upon UHMSF knockdown.**

A   Cassette exon allowing robust quantification of splicing in qPCR assays was selected from the NCBI Reference Sequences library. Then, each exon was scored for the presence of short polypyrimidine stretches (SPY) in the 5′ flanking region and the difference in this score with that for the downstream exon was calculated (see Materials and Methods and Appendix Fig S6). Similar numbers of cassette exons with high and low SPY scores were selected.

B   For the 65 selected cassette exons, the effect of knockdown of splicing factors was compared with the nonsilencing (NS) and scored by calculating first an absolute value for expression of the inclusion "$S_{in}$" and exclusion "$S_{out}$" isoforms and then Splicing Index and Splicing Changes with the indicated formulas (SIc: mean Splicing index for the control shRNA). Contrasting examples of results for two cassette exons are presented. The TJP1 exon 20 inclusion decreases upon U2AF[65] or CAPERα depletion, while the DNM2 exon 14 shows the opposite responses (mean ± SD for triplicate measures for two different shRNAs).

C   Comparison of Splicing Changes of 65 cassette exons upon knockdown of four UHMSFs.

D   Venn diagram for exons showing decreased inclusion upon knockdown, illustrating the functional relationship between U2AF[65] and CAPERα.

E   Correlation analysis of all Splicing Changes for the different UHMSF knockdowns.

F   Correlation between the effects of the knockdowns of the UHMSFs in HeLa cells and the ΔSPY_score of 65 cassette exons.

G   Correlation between the effects of UHMSF knockdowns in K562 cells and the ΔSPY_score of 5,629 cassette exons (RNA-Seq data analysis).

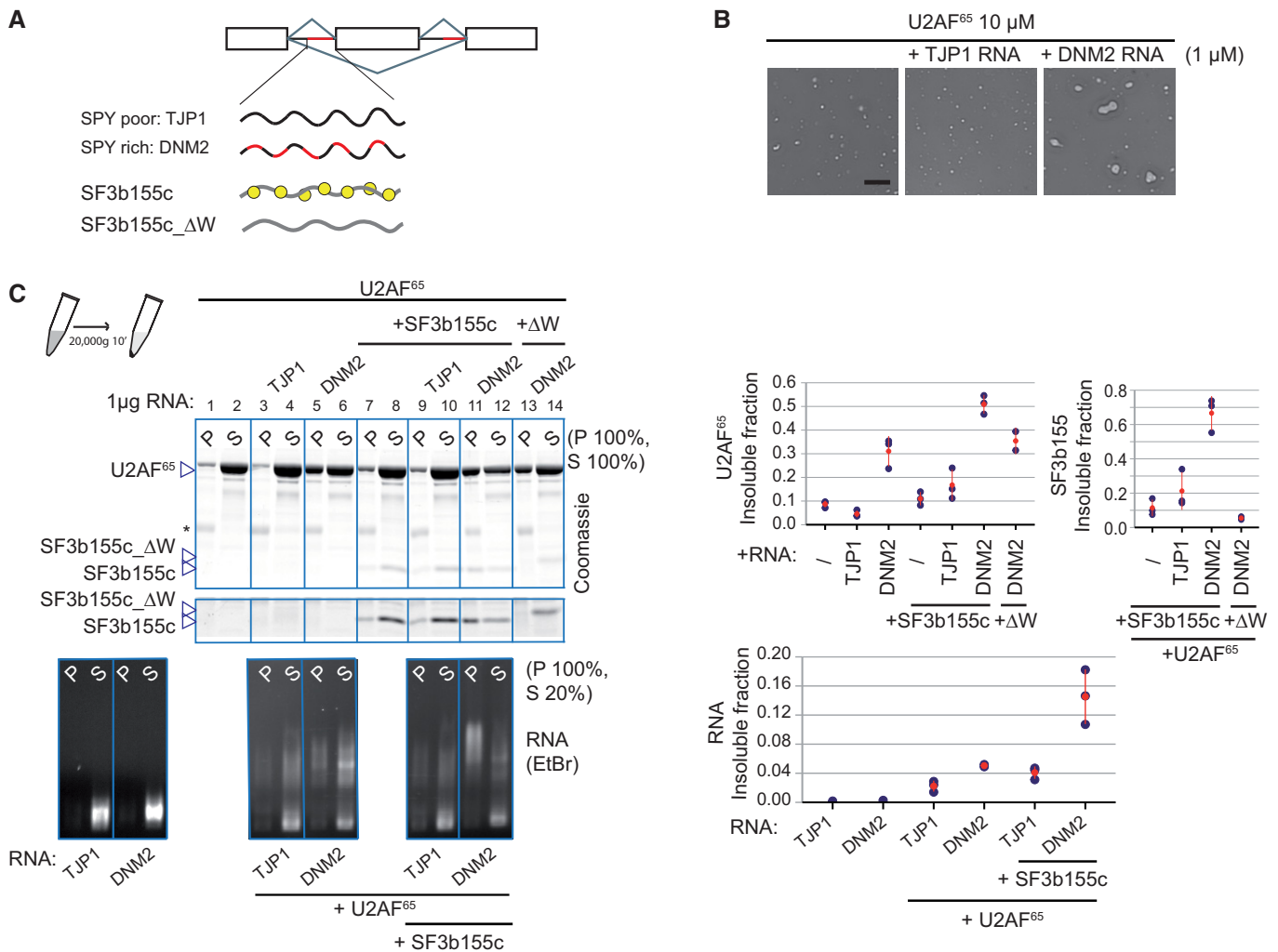Data information: Statistics: Spearman rho coefficient and null hypothesis t-test.

**Figure 5. U2AF[65] assemblies recruit RNA and SF3b155.**

A  Schematics of the RNA transcribed from intronic regions for experiments in (B) and (C) (see Appendix Fig S11 for details) as well as wild-type and mutant SF3b155 ULM-containing domains.

B  Analysis of U2AF[65] LLPS as in Fig 3, alone or in the presence of the SPY-poor TJP1 RNA or the SPY-rich DNM2 RNA. Representative phase-contrast images of two experiments. Scale bar: 10 μm.

C  Sedimentation assays of U2AF[65] (2 μM) in the presence of TJP1 or DNM2 RNAs (1 μg), and SF3b155c or SF3b155c_ΔW (1 μM) in Hepes 20 mM pH 7.5, 150 mM NaCl. The protein content in pellet and supernatant was analyzed by SDS–PAGE and Coomassie blue staining. The asterisk indicates a protein from skim milk used to block the tube surface before the assay. The RNA content was analyzed in parallel experiments by solubilizing the pellet with 0.1% SDS before agarose gel electrophoresis and ethidium bromide staining. Quantification of three experiments is shown (individual measures in blue and mean with SD in red).

recruitment of the SPY-rich RNA to U2AF[65] assemblies. Using mutants of SF3b155c with different combinations of ULM tryptophan residues, we observed that the two best tryptophan ligands of U2AF[65] were not sufficient for SF3b155c to enhance U2AF[65] sedimentation (Appendix Fig S19). In contrast, further addition of a lower affinity site conferred this property to SF3b155c. This effect of SF3b155c on U2AF[65] assemblies was not observed for SF1c which presents a single ULM (Appendix Fig S20). Altogether, the multiplicity of ULM domains is important for SF3b155 to interact with U2AF[65] assemblies and to enhance their formation or stability. We hypothesize that, similar to repeated SPY sequences on RNA, the multi-ULM domain of SF3b155 increases local concentration of U2AF[65], thereby favoring liquid-like U2AF[65] assemblies. Finally,

CAPERα addition to the sedimentation assays showed that combination with U2AF[65] and SF3b155c formed assemblies capable of recruiting the DNM2 intronic RNA with strong synergies (Fig EV3).

## The RS domain of U2AF[65] drives its speckled pattern and colocalization with SF3b155 in HeLa cells

To provide insights onto U2AF[65] assemblies *in vivo*, we analyzed the sedimentation properties of U2AF[65] from HeLa cell extracts and its nuclear segregation in speckles which are compartments enriched in splicing factors and other RNA-processing factors, and thought to depend on LLPS [47]. First, we adapted our
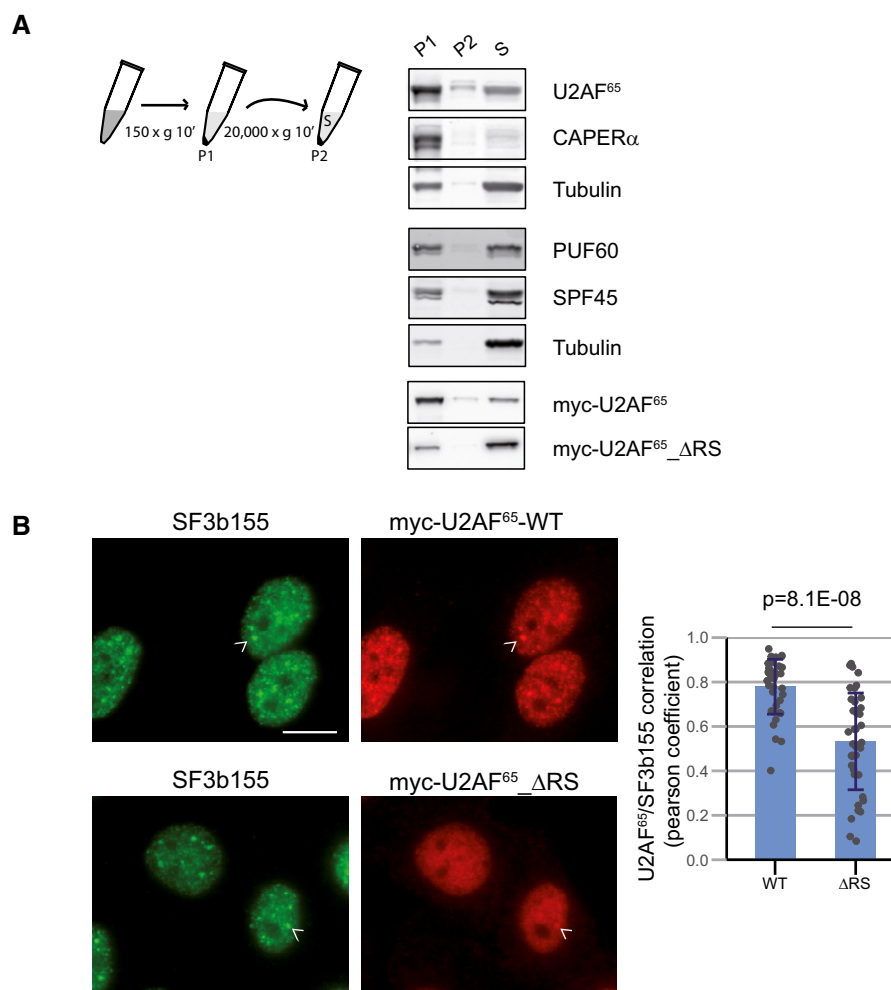
**A**



**B**



**Figure 6.  U2AF$^{65}$ sedimentation and localization to speckles in HeLa cells.**

A   HeLa cell extracts were prepared by sonication in 20 mM Hepes and 100 mM NaCl and analyzed by successive sedimentations and Western blot.
B   The nuclear localizations of myc-U2AF$^{65}$ and myc-U2AF$^{65}$_ΔRS were analyzed by immunofluorescence microscopy (scale bar: 2 μm). Arrowheads point to speckled labeling. Colocalization of U2AF$^{65}$ with SF3b155 was assessed by measuring the Pearson coefficient for the correlation of signals intensities along lines across the nuclei of 40 cells in each conditions (mean ± SD, statistics: Wilcoxon rank-sum test; representative results of three experiments).

sedimentation assay of recombinant proteins, to protein extracts of HeLa cells. HeLa cells were disrupted by sonication, and after DNase and RNase treatment, solubility was analyzed by successive low-speed and high-speed sedimentations (Fig 6A). Similarly to the recombinant proteins, U2AF$^{65}$ and CAPERα from HeLa cells were relatively deprived in the soluble fraction compared with PUF60 and SPF45. Overexpressed myc-tagged U2AF$^{65}$ presented a distribution similar to the endogenous protein, but deletion of the RS domain dramatically reduced its sedimentation. Immunofluorescence microscopy revealed a speckled repartition of myc-U2AF$^{65}$ in the nucleus as expected [48] (Fig 6B). This enrichment in speckles could also be detected for the RS domain-lacking protein as previously described [49], but the distribution we observed was more diffuse and the colocalization with SF3b155 was significantly reduced. Therefore, the U2AF$^{65}$ RS domain mediates multivalent interactions *in vitro* and localization to compartments thought to originate from LLPS *in vivo*.

## Discussion

### Correlation between UHMSFs in splicing regulations

In our knockdown approach in HeLa cells, we observed a general function of U2AF$^{65}$, CAPERα, and PUF60 to promote exon inclusion. In addition, correlation analyses pointed to a particular functional relationship between U2AF$^{65}$ and CAPERα in controlling exon cassette splicing, in agreement with a recent comparison of these splicing factors in MCF-7 cells [42]. An extensive analysis of splicing using RNA-Seq data in K562 cells confirmed the highest positive correlation between U2AF$^{65}$ and CAPERα and allowed to detect also a positive correlation of these two factors with PUF60 in agreement with redundancy and cooperativity of U2AF$^{65}$ and PUF60 for splicing *in vitro* [41].

In contrast with U2AF$^{65}$, CAPERα, and PUF60, we noted that SPF45 presented a general inhibitory action on cassette exon

inclusion. These results extend the previous observations that SPF45 inhibits FAS exon 6 and SXL exon 3 inclusion [24,50,51].

## Model of U2AF[65] and CAPERα liquid-like assemblies for 3′ splice site recognition

Using independent RNA-Seq data to complement our knockdown experiments in HeLa cells, statistical analyses clearly revealed a positive correlation between CAPERα and U2AF[65] decreased levels and the inclusion of cassette exons displaying repeated pyrimidine tracts in their 5′ flanking region. Other reports have documented similar observations: Fu and colleagues have proposed that U2AF[65] could have a long distance negative effect on spliceosome assembly by an unknown mechanism [44]. Vorechovski and colleagues suggested that U2AF[65] or CAPERα could facilitate the recruitment of inhibitory pyrimidine-binding proteins on long AG exclusions zones [45,52]. However, U2AF[65] is more known to be essential for spliceosome assembly [3,6,7,14]. If we consider that in a number of situations the cassette exon and the downstream exon compete for being first spliced, the increased inclusion of cassette exons can be interpreted as an improvement of the relative recognition of SPY-rich over SPY-poor 5′ flanking regions when U2AF65 or CAPERα levels are reduced (Fig 7A). Then, two scenarios can be envisaged: (i) Isolated U2AF[65] or CAPERα may display a higher affinity for pyrimidine-rich region due to the multiplicity of binding sites (avidity). As a result, an increased inclusion of exons with long pyrimidine-rich 5′ regions when competing with cassette exons with pyrimidine-poor 5′ regions should occur, in agreement with our results (Fig 7B, left). (ii) U2AF[65] and CAPERα may form liquid-like condensates on long pyrimidine-rich regions 5′ to cassette exons. A hallmark of LLPS is the critical concentration of proteins with self-attracting LCD above which liquid droplets are assembled. As long as the concentration of proteins harboring LCD is kept higher than the critical concentration, liquid droplets will be present. Decreasing their concentration below the critical concentration would lead to an abrupt dissociation of liquid droplets. If U2AF[65] or CAPERα liquid-like condensates are stabilized on long pyrimidine-rich regions, the critical concentration is lower in this specific environment. We might then observe the preferential dissociation of U2AF[65] and CAPERα from exons with pyrimidine-poor 5′ flanking regions (Fig 7B, right).

The in-depth biochemical analysis undertaken in this study strongly supports this second scenario. First, pulldown experiments show that heterotypic complexes of CAPERα and U2AF[65] are formed at the surface of the multi-ULM domain of SF3b155 and the multiplicity of ULMs is mandatory for this process. The RS domain of U2AF[65] is crucial for its efficient recruitment to SF3b155. Then, the RS domain of U2AF[65] forms multivalent interactions as evidenced by the observation of liquid-like droplets (Fig 3). SPY-rich RNAs but not SPY-poor RNAs reinforce the formation of U2AF[65] and CAPERα assemblies observed in sedimentation experiments or by microscopy. The multi-ULM domain of SF3b155 can reinforce the formation of U2AF[65] and CAPERα assemblies. Finally, the cosedimentation of SF3b155 and the preferential cosedimentation of SPY-rich over SPY-poor RNAs with U2AF[65] assemblies suggest that U2AF[65] liquid droplets could preferentially bridge U2 snRNP and pyrimidine-rich target RNAs. Altogether, our biochemical analyses clearly point toward liquid-like CAPERα or U2AF[65] assemblies being

responsible for a differential recognition of repeated pyrimidine tracts to explain the *in vivo* splicing results (Fig 7C).

In contrast with U2AF[65] and CAPERα, our analysis of RNA-Seq data point to a negative relationship of Splicing Changes upon PUF60 knockdown with ΔSPY scores in HEK293T cells and a less pronounced one in K562 cells (Fig EV2C). Similarly, Vorechovsky and colleagues observed opposite effects of U2AF and PUF60 knockdown on cassette exons preceded by extended uridine-rich regions in HEK293T cells [45]. Further investigations are needed to state the significance of this overlap between PUF60-repressed and U2AF[65]-activated exons and its molecular basis. Of note, global analysis showed a positive correlation of Splicing Changes upon U2AF[65] and PUF60 depletion in K562 cells (Fig EV2B), suggesting generally similar and possibly concerted actions of these factors in exon inclusion.

U2AF[65] and CAPERα share a similar primary structure and the property to form large assemblies, but a significant number of splicing events are differently affected by their knockdown. Therefore, CAPERα adds the possibility to regulate a set of exons with a partial overlap with those regulated by U2AF[65] abundance. Deciphering whether specific RNA codes are recognized by U2AF[65] and CAPERα assemblies, and what are the intervening actions of PUF60, SPF45, and tat-SF1 in these processes should provide a deeper understanding of the complex recognition of splice sites for correct reading of the genomic information.

Recent data show that the low complexity sequences of splicing factors RBFOX1 and hnRNPD are responsible for formation of large assemblies in which they may exert their action in splicing [53,54]. Our data add to this list of splicing factors that can form dynamic networks of interactions and importantly also provide evidences that repeat in RNA sequences can control liquid-like behavior of splicing factors. Interestingly, Gladfelter and colleagues reported that two mRNAs could differently affect LLPS by Whi3, an RNA-binding protein with a polyQ expansion [55]. In addition, Hyman and colleagues reported that competition for mRNA controls PGL-3 LLPS [56]. Therefore, interplays between RNA and low complexity domain proteins appear to be important in LLPS.

## RS domain-dependent formation of macromolecular assemblies

RS domains are known to mediate both protein–RNA and protein–protein interactions (reviewed in ref. [57]). Interactions between RS domains are proposed to function in bridging proteins bound to exonic splicing enhancers with spliceosomal components recruited to splice sites, thus helping exon definition. RS domains are also involved in the routing of splicing factors to nuclear speckles which are membrane-less structures enriched in splicing and other RNA-processing factors that, as nucleoli or Cajal bodies, are thought to self-organize by mean of LLPS [47,58,59]. The RS domain of U2AF[65] is more known to be implicated in contacting and facilitating the formation of the U2 snRNA-pre-mRNA duplex [14,60].

RS domains can confer aggregation properties to proteins [29–31]. Here, we provide evidence that the RS domain of U2AF[65] drives LLPS *in vitro*, suggesting the low complexity nature of this domain is a molecular basis for multivalent protein–protein interactions. This result suggests that, more generally, assemblies of RS domain-containing proteins can be dynamical structures with important functional implications for splicing.
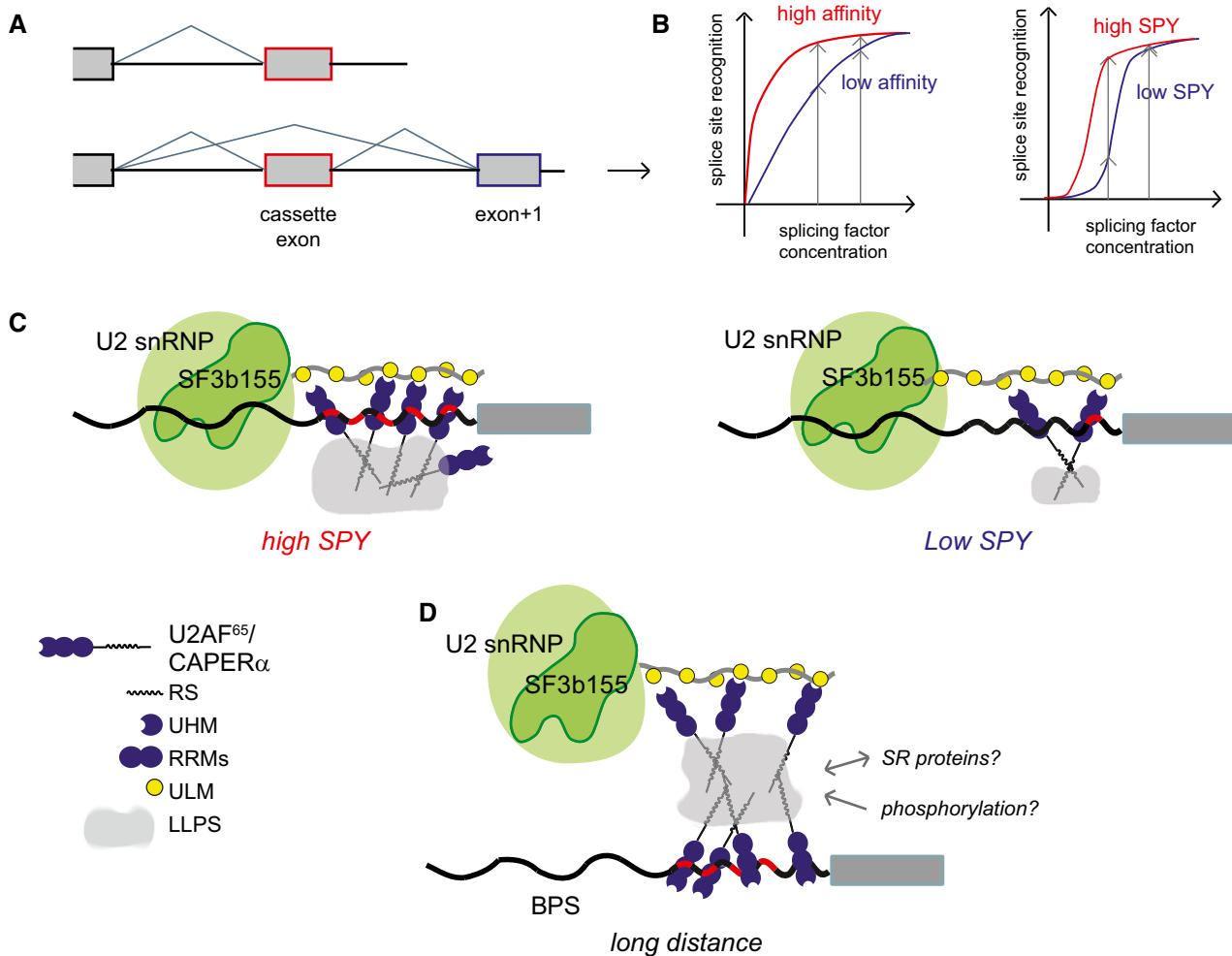
**Figure 7. Role of large assemblies of U2AF[65] and CAPERα in splice site recognition.**

A   Splicing of cassette exons implies a competition for recruitment of splicing factors at 3′ splice sites.

B   Competition between splice sites determines cassette exon inclusion. The reduction in a splicing factor concentration can lead to increased inclusion of a cassette exon in particular contexts: left: target sequences with different affinities, right: sigmoid curve responses due to target sequences recognition by cooperative assemblies of splicing factors.

C   Model for different geometries of 3′ splice site recognition by U2AF[65] and CAPERα assemblies. Repeated polypyrimidine tracts (PPTs) stabilize SF3b155 at the branch site and can favor splice site recognition upon U2AF[65] or CAPERα reduction.

D   RS domain-mediated LLPS could mediate long-range interactions for splice site recognition and provide a mean for regulation by kinases and phosphatases or interaction with other splicing factors.

Carmo-Fonseca and colleagues have shown that the deletion of the RS domain of U2AF[65] does not prevent its accumulation in speckles but impairs its targeting to sites of active splicing [49]. We also observed that U2AF[65] deprived of its RS domain could still accumulate in nuclear speckles but with a reduced efficiency. Therefore, multivalent interactions mediated by its RS domain probably contribute together to U2AF[65] local concentrations in the nucleus and to the formation of macromolecular assemblies contacting the multi-ULM domain of SF3b155 and polypyrimidine-rich sequences in RNA.

Extensive phosphorylation of RS domains influences the interactions, the subcellular localization, and the functions of splicing factors [57]. In our experiments, we used U2AF[65] and CAPERα from cell extracts or recombinant U2AF[65] produced in bacteria in the presence of the kinase SRPK1, tentatively mimicking the phosphorylation state of the endogenous protein [61]. Further investigations will address how phosphorylations of their RS domains regulate the size and dynamics of U2AF[65] and CAPERα assemblies, thus providing a control for U2 snRNP recruitment (Fig 7D).

Interactions with additional partners could also regulate RS domain-dependent U2AF[65] and CAPERα assemblies. In particular, U2AF[65] is generally thought to be tightly associated with U2AF[35] to form a U2AF heterodimer. U2AF[35] also possesses an RS domain. As it was shown that the presence of at least one of the U2AF RS domains is necessary for drosophila viability, we envisage that the RS domain of U2AF[35] can participate in the formation of U2AF[65] assemblies [62].

## Putative specific functions of U2AF$^{65}$ and CAPERα assemblies

In addition to the preferential recognition of SPY-rich RNA for splice site choice, the formation of macromolecular assemblies by U2AF$^{65}$ and CAPERα could bring distinct properties to the splicing machinery:

At a nanometric scale, U2AF$^{65}$ and CAPERα assemblies could trigger an initial long-range interaction between specific RNA sequences in a pre-mRNA and components of the splicing machinery (Fig 7D). As such assemblies are dynamic and can vary in size and shape, they could also provide a malleable surface to accommodate diverse distances between branchpoints and 3′ splice sites and changing spatial organizations of the pre-mRNA and splicing factors during early spliceosome assembly. U2AF$^{65}$ assemblies are also good candidates to contact SR proteins bound to exonic splicing enhancers, thus facilitating splice site recognition. The C-terminal domain (CTD) of polymerase II has been shown to bind polymers of LCD proteins [63]; as other studies reported the interaction of the CTD with U2AF$^{65}$ [64–66], we can hypothesize that U2AF$^{65}$ assemblies play a role in bridging the splicing and the transcription machineries. LLPS of U2AF$^{65}$ could also reduce diffusion from transcription sites, as recently proposed more generally for splicing factors harboring LCD [67]. This is supported by our results and the observation by Carmo-Fonseca and colleagues that the RS domain is required for U2AF$^{65}$ accumulation at transcription sites [49].

# Materials and Methods

### Antibodies

The following antibodies were used:
Bethyl: PUF60 rabbit polyclonal Ab, A302-818A; CAPERα rabbit polyclonal Ab, A300-291A; SPF45 rabbit polyclonal Ab, A302-497A; U2AF$^{65}$ rabbit polyclonal Ab, A303-667A; SF3b155 rabbit polyclonal Ab, A300-997A.
Sigma: U2AF$^{65}$ mouse monoclonal Ab, clone MC3. Mouse monoclonal Anti-myc clone 9E10.
Santa Cruz: PUF60 mouse monoclonal Ab, sc-398785; CAPERα mouse monoclonal Ab, sc-101103.
Novusbio: anti-FUS rabbit polyclonal Ab, NB100-565.

SC35 mouse monoclonal supernatant was a kind gift of John Bell. Tubulin alpha E7 hybridoma supernatant was produced in house.

### Proximity ligation assays

A proximity ligation assay kit from Olink Bioscience was used. HeLa cells were plated at 10,000 cells/wells on glass coverslips. After 48 h, cells were fixed with a PBS-paraformaldehyde 4% solution for 20 min at 37°C and permeabilized for 5 min at 37°C with 0.1% Triton X-100 in PBS. After two PBS washes, a blocking solution (Duolink II Fluorescence, Olink Bioscience) was added for 20 min at room temperature in a humid chamber. Each coverslip was then incubated overnight at 4°C with 30 μl of a mix of two antibodies (a mouse and a rabbit primary antibody) recognizing the two proteins of interest in a solution of PBS, Triton 0.1%, and BSA 3%. After two PBS washes, the coverslips were incubated first for 1 h at 37°C with

a mix of PLA® plus mouse and PLA® minus rabbit probes, and then with a mix composed of a ligation solution 1/5 (Duolink II Fluorescence, Olink Bioscience) and a ligase 1/40 for 30 min at 37°C. After two PBS washes, an amplification solution 1/5 (Duolink II Fluorescence, Olink Bioscience) and polymerase 1/80 were added on the coverslips for 100 min at 37°C. The nucleus was stained with DAPI, and the coverslips were mounted on slides using a mounting medium (Mowiol). The fluorescence was observed with a microscope Leica DM4 B using the Leica Las X software.

### GST-pulldown experiments

Cells grown in 35-mm-diameter wells were washed twice in PBS and lysed by vortexing in 1 ml of buffer A (50 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1% nonidet P40, 1 mM DTT, 10 μg/ml RNase A, and antiprotease mix from Roche Diagnostics). Cell extracts were clarified by centrifugation at $20,000 \times g$ for 10 min. RNA digestion was monitored by agarose gel electrophoresis and ethidium bromide staining. For each pulldown experiment, 25–40 pmol of purified GST-fusion protein was used as bait and equal volumes of clarified cell extract and interaction buffer I (10 mM Tris–HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 0.05% NP40). The interaction reactions were incubated for 1 h at 4°C. Glutathione beads (10 μl; GE Healthcare) were washed twice with buffer I, incubated with the interaction reaction for 30 min, washed four times with buffer I, and the retained proteins were separated by SDS–PAGE and revealed by immunoblot with 700 nm or 800 nm IRDye-conjugated antibodies (Rockland Immunochemicals). The fluorescence signal was acquired with an infrared laser scanner (Odyssey, LI-COR, or Typhoon from GE Healthcare). Quantification was performed with ImageJ.

### Recombinant protein purification

Constructs in pGEX6p1 were cloned in BL21 or the derived C43 bacteria cell line. RS domain-containing proteins were coexpressed with SRPK1 to improve the yield of purified protein as described for SRSF1 [61]. Cell culture (1 l) in LB medium was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h at 30°C, and cells were harvested and frozen at −80°C. After thawing, the pellets were sonicated for 10 min at 30% power with a sonicator (Vibracell, Bioblock Scientific) in buffer containing 25 mM Hepes pH 7.5, 1 M NaCl, 1 mM EDTA, 10% glycerol (v/v), 5 mM DTT, 0.1 mM PMSF (phenylmethylsulfonyl fluoride), and protease inhibitors (cOmplete, Roche Diagnostic). The lysate was clarified by centrifugation for 40 min at $15,000 \times g$, and the soluble supernatant was loaded on a 1 ml GST-trap column (GE Healthcare) using a ÄKTA start chromatography system (GE Healthcare). Elution was performed with 40 mM reduced glutathione in 25 mM Hepes pH 7.5 and 1 M NaCl, and concentrated eluates were loaded on a Superdex 200 increase 10/300 column for gel filtration refinement in 25 mM Hepes pH 7.5 and 1 M NaCl. Purified proteins were concentrated, and salt concentration was reduced to 200 mM by centrifugal filtration. Protein concentrations were estimated by UV 280 nm light absorption and further determined by SDS–PAGE Coomassie blue staining against a BSA (bovine serum albumin) standard (Pierce). Proteins were snap-frozen with liquid nitrogen and kept at −80°C until used.

## Bioinformatic selection of cassette exon splicing events

Using a Galaxy pipeline (on the server http://use.galaxy.org and a local instance [68]), the collections of human (hg38) and mouse (mm10) exons of the NCBI Reference Sequence (RefSeq) library [39] were downloaded using the genome browser of the University of California of Santa Cruz (UCSC). Exons were filtered for the absence of alternative 5′ and 3′ splice site, for the existence of unique 5′ splice site for the preceding and 3′ splice site for the following intron in all annotated transcripts, and finally for the exons being present in at least one and absent in at least one of the annotated transcripts as depicted in Appendix Fig S6. Conserved cassette exon in human and mouse was retained by joining chromosome coordinates after converting mouse coordinates to human. The corresponding exon–exon junctions were then used as a library for Bowtie alignment of published RNA-Seq sequences for HeLa cells in standard conditions (~ 160 M reads) [69]. Splicing events with at least two reads aligned on both the skipped and inclusion isoforms and with a sum of at least 20 reads aligning on both isoforms were retained. For the selected exons, the 100 base 5′ flanking region was scored for the enrichment in putative U2AF[65]-binding sites (SPY score, see below) and similar numbers of low SPY score and high SPY score cassette exons were selected (red squares in Appendix Fig S6).

For the 65 selected cassette exons, primers specific for the "spliced in" ($S_{in}$) and "spliced out" ($S_{out}$) isoforms were designed using primer3 and checked for amplification efficiency and specificity by qPCR on HeLa total RNAs (Appendix Table S2).

## Scoring of small polypyrimidine tracts (SPY)

The heuristic scoring method which was used in this study was derived from a scoring method for polypyrimidine tract [70] and also based of structural features of the interaction of U2AF[65] RRMs with pyrimidine-rich sequences [38,40].

Small polypyrimidine tract (SPY) representing potential U2AF[65]-binding sites in the 100-bp 5′ flanking sequence of cassette exons was identified by scanning using excel formulas to find all sequences of eight or nine bases with the following characteristics.

1  Both 3′ and 5′ ends must be pyrimidines;
2  No contiguous purines are allowed;
3  Every purine must be surrounded by at least four pyrimidines;
4  Overlapping SPY sequences are not allowed with selection based on maximizing the global SPY score (see below).

For every detected SPY, a score was calculated based on the sequence length and content according to the following formula:

$$\text{score}(\text{SPY}) = \sum_{x \in \{G,A,T,C\}} f(x) \cdot s(x)$$

Where $f(x)$ is the occurrence of nucleotide $x$ in the SPY and $s(T) = 3$, $s(C) = 2$, $s(A) = -2$, and $s(G) = -2$.

Each 100 base FIR 5′ to cassette exon was then given a global SPY score (SPY score) as the sum of scores for all SPY detected in this region. In case of overlap of several SPYs, the combination of SPYs yielding the maximal global SPY score was retained. We also calculated the difference between the SPY score restricted to AG exclusion zones for the cassette exon and the following exon (ΔSPY_score used in Figs 4 and EV2).

## Cell culture, shRNA transduction

HeLa cells are maintained in culture in DMEM medium supplemented with 10% fetal bovine serum at 37°C with 5% $CO_2$. The day before transduction, 100,000 cells were plated in 3.5-cm-diameter wells of a six-well plate with 3 ml of cell culture medium. In order to reduce the levels of CAPERα, U2AF[65], PUF60, or SPF45, the cells were infected using 40 μl of viral supernatant for shRNA expression diluted in 1 ml of cell culture medium with 8 μg/ml of polybrene. Two shRNAs were used for each gene (Appendix Table S1), and each knockdown was performed in triplicate (six samples for each knockdown). As control, HeLa cells were transduced with a nonsilencing viral supernatant.

The day following infection, 1 μg/ml of puromycin was added to the cell medium; selection was maintained for 72 h, and cells were harvested for RNA purification. At that time point, all cells in the control nontransduced well had died.

## RNA isolation and reverse transcription

The transduced cells were washed with PBS and total RNAs were purified using a NucleoSpin® RNA kit according to the supplier's recommendations (Macherey-Nagel) or TRIpure reagent (Euromedex). Precipitated RNAs were resuspended in sterile water, and the concentration was determined by spectrophotometry.

1 μg of total RNAs was hybridized with 1 μg of a random sequence hexamer oligodeoxyribonucleotides for 5 min at 65°C and then put on ice for 5 min. The reverse transcription reaction was performed in a 20 μl volume containing 200 units of reverse transcriptase ImProm-II™ (Promega), 20 units of RNasin (Euromedex), and 10 mM of each nucleotide in the specific enzyme buffer.

### High-throughput qPCR analyses (Biomark HD, Fluidigm)

#### Preamplification of cDNAs
For Biomark HD analyses, cDNAs were first preamplified using a mix of all oligonucleotides to be later used in individual qPCRs (each primer at a concentration of 500 nM in 10 mM Tris–Cl, pH 8.0, 0.1 mM EDTA). Preamplification reactions were performed in a 5 μl volume containing 1.25 μl of cDNA, 1 μl of preamplification mix PreAmp Master Mix (Fluidigm PN 100-5580), 0.5 μl of the oligonucleotides mix, and 2.25 μl sterile water. Ten PCR cycles were performed (15 s at 95°C, 4 min at 60°C) using an Applied Biosystems 9700 thermocycler. To digest unincorporated primers, reaction products were treated for 30 min at 37°C with 2 μl of Exonuclease I (4 units/μl; New England Biolabs, PN MO0293S). Samples were then incubated for 15 min at 80°C to inactivate the nuclease. Then, 7 μl final product was finally diluted with 43 μl of suspension buffer (10 mM Tris–Cl pH 8.0, 0.1 mM EDTA).

#### Measure of gene expression with 48.48 integrated fluidic circuits
Microfluidic arrays 48.48 integrated fluidic circuits (IFC; Fluidigm) were primed with the Fluidigm IFC Controller. Then, a total volume of 5 μl per sample containing 2.25 μl of preamplified cDNA, 2.5 μl

of supermix (Bio-Rad PN 172-5211), and 0.25 µl of DNA-binding dye (Fluidigm PN 100-7609) was distributed in each sample inlets of the IFC. For each assay, a 5 µl mix containing 2.5 µl of assay loading reagent (Fluidigm PN 100-7611), 0.25 µl of a mix of forward and reverse oligonucleotides (each at 100 µM), and 2.25 µl of suspension buffer was distributed in each assay inlets. Samples and oligonucleotides were then loaded in the microfluidic array capillaries using the Fluidigm IFC Controller.

PCR was then performed using the Biomark HD and the Data Collection software. CT values were calculated with the Fluidigm real-time qPCR analysis software, and expression was calculated from the $\Delta$CT using GAPDH for normalization.

### RNA-seq data analyses

The following data were used for our analyses:

K562 (ENCODE consortium): U2AF2 knockdowns SRR3469570 and SRR3469571 and paired controls SRR3469464 and SRR3464465; CAP ERα knockdowns SRR442211 and SRR4422113 and paired controls SRR4421418 and SRR4421417; PUF60 knockdowns SRR4421954 and SRR4421955 and paired controls SRR4422045 and SRR4422044; and SPF45 knockdowns SRR4422466 and SRR4422465 and paired controls SRR4421359 and SRR4421360.
MCF7: RBM39 knockdown SRR4239823, U2AF2 knockdown SRR4 239824, and control SRR4239822.
293T cells: controls ERR2223563 and ERR2223564, RBM39 knockdowns ERR2223565 and ERR2223566, and PUF60 knockdowns ERR2223567 and ERR2223568.
HeLa cells: U2AF2 knockdown SRR1582594 and control SRR1582593.

For splicing analyses, reads were aligned to our collection of exon junctions with Bowtie using the default parameters but a maximum edit distance of 2. Numbers of aligned reads were counted using a Galaxy workflow (http://use.galaxy.org and http://use.galaxy.eu), and the results were then further analyzed in a calc spreadsheet. Splicing Changes were calculated as for qPCR data (see previous paragraph). A filter was applied to remove splice junctions with few aligned reads: Events were removed when a single-read difference would lead to a variation of the Splicing Change above 0.02. Remaining cassette exons were used for correlation analyses by calculating the Pearson coefficient on the ranked values.

### Splicing Index and splicing Change calculation

Based on qPCR CT values or numbers of aligned reads for RNA-Seq data, an absolute expression value was attributed to both exon inclusion and exclusion isoforms ($S_{in}$ and $S_{out}$). The splicing Index belonging to [0,1] was calculated as usual:

$$SI = \frac{S_{in}}{(S_{in} + S_{out})}$$

To measure the effect of knockdown on splicing, we introduced a splicing Change index that corresponds to the ratio of the splicing Index difference to the maximal theoretical increase. This index is therefore a value belonging to [−1,+1].

$$SC = \frac{(SI - SI_c)}{(1 - MIN(SI, SI_c))}$$

### Statistics

Quantification values of coomassie or immunoblots signals for independent replicates are presented as means ± SEM with each individual experiment indicated. Excel or Rstudio was used for statistical analyses. Student's *t*-test (unpaired, two-sided *t*-test) or nonparametric Wilcoxon test was applied when indicated. For correlation analyses of splicing, we calculated Spearman's correlation rho coefficient as the Pearson coefficient for the rank of Splicing Change values for each knockdown and the rank of $\Delta$SPY_scores for each cassette exon. The null hypothesis of whether Spearman's rho was significantly different from zero was then tested *via t*-test.

### Liquid phase analyses

Concentrated proteins were diluted twice to reach the appropriate final salt concentration. 2 µl of this dilution was loaded in-between slide and coverslip and analyzed using an Eclipse Ti microscope with Nikon S Plan fluor 40 × 0.6 objective. In some instance, a modification of the observed pattern was noticed over time and was probably dependent on initial stretching of the droplets during coverslip deposit. Video microscopy was performed to document this phenomenon using a Neo sCMOS camera (Andor Technology) and NIS-element imaging software.

### Recombinant protein sedimentation assays

After thawing, recombinant proteins were centrifuged at $20,000 \times g$ for 10 min and the soluble part was used to assemble the sedimentation assays in 10 µl in 20 mM Hepes pH 7.5 and indicated salt concentrations using low-binding Eppendorf 1.5-ml microtubes. The tube surface was previously coated by incubation with skim milk (1% in 20 mM Tris–Cl pH 7.5 and 100 mM NaCl) for 5 min to avoid nonspecific absorption of proteins during the assay. After incubation for 1 h of the proteins and RNA mixes at 4°C, tubes were centrifuged at $20,000 \times g$ for 10 min and the supernatant was transferred to a new tube and adjusted at 0.1% SDS and 3% glycerol (v/v) with xylene cyanol traces. The pellets were resuspended using 20 mM Tris–Cl, 100 mM NaCl, 0.1% SDS (to help dissociating the pellet), 3% glycerol (v/v), and traces of xylene cyanol. The indicated fractions of the pellets and supernatants were loaded on 1% agarose gel in 0.5× TAE buffer and analyzed by SDS–PAGE. In these conditions, RNA migration was not reproducible probably due to variable denaturation of the proteins by SDS, but the amount of RNA could be quantified. SDS–PAGE gels were stained with Coomassie blue and quantified using an Odyssey scanner (LI-COR) by scanning with an excitation wavelength of 685 nm.

### RNA *in vitro* transcription

Substrate for *in vitro* transcription was prepared by amplification of a 130 base-pair DNA fragment using genomic DNA from HeLa cells as template and the following primers:

TJP1_if TAATACGACTCACTATAGGGTTAGGACATGAGGTCATAATTTTGT

TJP1_ir CTGTGAAGTGTTTAAAATATTTTAAATAT

DNM2_if TAATACGACTCACTATAGGGCCTGCTCTTCTCTTCTCCTGT

DNM2_ir CTTTGAATATGCAAAGTTAGGGTTGT

PPHLN1_if TAATACGACTCACTATAGGGAAAGAGTTTAGGGCCACCAGA

PPHLN1_ir CTATGGAAAGGTGGGCAGGG

ZNF207_if TAATACGACTCACTATAGGGAATGAATATGAGGTGCTTGTGCC

ZNF207_ir CTGTGAAAATGTATGACAAAATCAAAATCT

TIA1_if TAATACGACTCACTATAGGGCCCAGTAGTTTTATTGTAAAGCCTAT

TIA1_ir CTGATGACAAAGATTAGATTTGTTCTTAA

TCF20_if TAATACGACTCACTATAGGGTTGTCGGTCAGGACTGACTC

TCF20_ir CTGACACGGGCAAAACCAAG

Purified DNA amplicons were then used for *in vitro* transcription with T7 RiboMAX kit from Promega. RNA was analyzed and quantified by dilution in a denaturing buffer and rapid migration on agarose gel (Appendix Fig S11).

### Gel shift analyses of RNA

*In vitro* transcribed RNA and purified U2AF$^{65}$ were mixed at different concentrations in 10 μl containing 20 mM Tris–Cl pH 7.5, 140 mM NaCl. After a 5-min incubation at 4°C, 1 μl of 10× buffer containing 250 mM Tris–Cl pH 7.5, 30% glycerol (v/v), and traces of xylene cyanol was added and the mixture was separated by quick migration in 1% agarose gel containing 10 μg/ml ethidium bromide and 0.5× Tris-acetate buffer and visualized under UV exposure. Bands were quantified using ImageJ.

### Minigene preparation

DNA fragments of interest were amplified by PCR from HeLa cell genomic DNA for the TJP1 sequence, or a synthetic DNA template for DNM2. For TJP1, the minigene sequence covers all genomic sequences starting with 24 bases at the 3′ end of exon −1 up to 24 bases at the 5′ end of exon +1. For DNM2, the minigene sequence covers 24 bases at the 3′ end of exon −1, 150 bases at each extremity of intron −1, the cassette exon, 150 bases at each extremity of intron +1, and 24 bases at the 5′ end of exon +1. These fragments were inserted by restriction-free cloning (www.rf-cloning.org) after the sixth base and in frame with the coding region of firefly luciferase in vector CMV-LUC2CP/ARE [71].

### Reporter assays

HeLa cells were infected with viral supernatants for shRNAs expression. At day 3, 5 ng of minigene DNA was transfected using Lipofectamine 2000 (Invitrogen). Cells were maintained in culture for 24 h before RNA purification. Splicing analysis was performed as described previously with modifications [72]. 10 μl qPCRs included 4 μl of a SYBR Green 2× Master Mix (GoTaq, Promega), 1 μl of a 2.5 μM specific primers mix, and 5 μl of diluted cDNA. The reaction

was performed with a thermocycler CFX384® (Bio-Rad), and RNA was quantified using the delta-CT method.

### Sedimentation of cell extracts

HEK293 in a 35-mm-diameter wells (80% confluence) was washed twice in PBS, and then resuspended in 200 μl of 20 mM Hepes pH 7.5, 100 mM NaCl, with protease inhibitors (cOmplete, Roche Diagnostic), MgCl$_2$ 5 mM, CaCl$_2$ 1 mM, DNase 0.1 μg/μl, and RNase 25 ng/μl. The mixture was sonicated on ice twice 10 s at power 40 (Bioblock Ultrasonic Processor 75038, Bioblock Scientific) with a 20-s interval. Examination under the microscope ascertained that no intact nuclei remained in these conditions. After 15 min at room temperature, tubes were centrifuged at 150 × *g* for 5 min, and the upper 100 μl was further centrifuged at 20,000 × *g* for 10 min. The first and second pellets (p1 and p2) were resuspended in Laemmli, and equivalent of p1, p2, and soluble fraction was analyzed by Western blotting. We checked in these conditions that the first supernatant was free of DNA and RNA by migration on ethidium bromide agarose gels.

### Immunofluorescence for analyses of the nuclear localization of U2AF$^{65}$

HeLa cells were transfected with pCDNA3-based expression plasmids for myc-U2AF$^{65}$ and myc-U2AF$^{65}$_ΔRS. After 3 days, cells were fixed in methanol 8′ and 4% paraformaldehyde (10′ at room temperature). Standard immunolabeling was performed with anti-myc (1/200) and anti-SF3b155 (1/200). Cells were analyzed with a Zeiss Axiovert 200 M and ×64 oil objective. For image acquisition, fields were selected randomly to avoid any trend. Images were analyzed using ImageJ software. Signals were measured along a line drawn through the nucleus avoiding nucleoli (as both proteins showed exclusion). Spearman correlation was determined for 40 nuclei in each condition, and the results were compared using Wilcoxon rank-sum test.

## Data availability

All relevant data are included in the main article and supplementary figures. Additional information including raw qPCR data is available from the corresponding author upon request.

**Expanded View** for this article is available online.

## References

1. Papasaikas P, Valcárcel J (2016) The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem Sci* 41: 33–45

2. Wahl MC, Will CL, Lührmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136: 701–718

3. Zamore PD, Green MR (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J* 10: 207–214

4. Zorio D, Blumenthal T (1999) Both subunits of U2AF recognize the 3′ splice site in *Caenorhabditis elegans*. *Nature* 402: 835–838

5. Wu S, Romfo CM, Nielsen TW, Green MR (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. *Nature* 402: 832–835

6. Ruskin B, Zamore PD, Green MR (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52: 207–219

7. Zamore PD, Green MR (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86: 9243–9247

8. Zamore PD, Patton J, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355: 609–614

9. Kielkopf CL, Lücke S, Green MR (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 18: 1513–1526

10. Loerch S, Kielkopf CL (2016) Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. *RNA* 22: 1795–1807

11. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Krämer A, Sattler M (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* 11: 965–976

12. Wang W, Maucuer A, Gupta A, Manceau V, Thickman KR, Bauer WJ, Kennedy SD, Wedekind JE, Green MR, Kielkopf CL (2013) Structure of phosphorylated SF1 bound to U2AF65 in an essential splicing factor complex. *Structure* 21: 197–208

13. Zhang Y, Madl T, Bagdiul I, Kern T, Kang H-SS, Zou P, Mäusbacher N, Sieber SA, Krämer A, Sattler M (2013) Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3′-splice site recognition. *Nucleic Acids Res* 41: 1343–1354

14. Valcárcel J, Gaur RK, Singh R, Green MR (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* 273: 1706–1709

15. Gozani O, Potashkin J, Reed R (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 18: 4752–4760

16. Abovich N, Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89: 403–412

17. Berglund J, Abovich N, Rosbash M (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* 12: 858–867

18. Rain J, Rafi Z, Rhani Z, Legrain P, Krämer A (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. *RNA* 4: 551–565

19. Manceau V, Swenson M, Le Caer J-P, Sobel A, Kielkopf CL, Maucuer A (2006) Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *FEBS J* 273: 577–587

20. Cass DM, Berglund J (2006) The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry* 45: 10092–10101

21. Thickman KR, Swenson MC, Kabogo JM, Gryczynski Z, Kielkopf CL (2006) Multiple U2AF65 binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *J Mol Biol* 356: 664–683

22. Manceau V, Kielkopf CL, Sobel A, Maucuer A (2008) Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *J Mol Biol* 381: 748–762

23. Loerch S, Maucuer A, Manceau V, Green MR, Kielkopf CL (2014) Cancer-relevant splicing factor CAPERα engages the essential splicing factor SF3b155 in a specific ternary complex. *J Biol Chem* 289: 17325–17337

24. Corsini L, Bonnal S, Bonna S, Basquin J, Hothorn M, Scheffzek K, Valcárcel J, Sattler M (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat Struct Mol Biol* 14: 620–629

25. Dowhan DH, Hong EP, Auboeuf D, Dennis AP, Wilson MM, Berget SM, O'Malley BW (2005) Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. *Mol Cell* 17: 429–439

26. Mercier I, Gonzales DM, Quann K, Pestell TG, Molchansky A, Sotgia F, Hulit J, Gandara R, Wang C, Pestell RG *et al* (2014) CAPER, a novel regulator of human breast cancer progression. *Cell Cycle* 13: 1256–1264

27. Huang G, Zhou Z, Wang H, Kleinerman ES (2012) CAPER-α alternative splicing regulates the expression of vascular endothelial growth factor165 in Ewing sarcoma cells. *Cancer* 118: 2106–2116

28. Wang E, Lu SX, Pastore A, Chen X, Imig J, Chun-Wei Lee S, Hockemeyer K, Ghebrechristos YE, Yoshimi A, Inoue D *et al* (2019) Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell* 35: 369–384

29. Nikolakaki E, Drosou V, Sanidas I, Peidis P, Papamarcaki T, Iakoucheva LM, Giannakouros T (2008) RNA association or phosphorylation of the RS domain prevents aggregation of RS domain-containing proteins. *Biochim Biophys Acta* 1780: 214–225

30. Krainer A, Mayeda A, Kozak D, Binns G (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell* 66: 383–394

31. Ge H, Zuo P, Manley J (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell* 66: 373–382

32. Wheeler RJ, Hyman AA (2018) Controlling compartmentalization by non-membrane-bound organelles. *Philos Trans R Soc Lond B Biol Sci* 373: 20170193

33. Courchaine EM, Lu A, Neugebauer KM (2016) Droplet organelles? *EMBO J* 35: 1603–1612

34. Ellis JD, Llères D, Denegri M, Lamond AI, Cáceres JF (2008) Spatial mapping of splicing factor complexes involved in exon and intron definition. *J Cell Biol* 181: 921–934

35. Loerch S, Leach JR, Horner SW, Maji D, Jenkins JL, Pulvino MJ, Kielkopf CL (2019) The pre-mRNA splicing and transcription factor Tat-SF1 is a functional partner of the spliceosome SF3b1 subunit via a U2AF homology motif interface. *J Biol Chem* 294: 2892–2902

36. Kielkopf CL, Rodionova NA, Green MR, Burley SK (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 106: 595–605

37. Stepanyuk GA, Serrano P, Peralta E, Farr CL, Axelrod HL, Geralt M, Das D, Chiu H-JJ, Jaroszewski L, Deacon AM *et al* (2016) UHM-ULM interactions in the RBM39-U2AF65 splicing-factor complex. *Acta Crystallogr D Struct Biol* 72: 497–511

38. Agrawal AA, Salsi E, Chatrikhi R, Henderson S, Jenkins JL, Green MR, Ermolenko DN, Kielkopf CL (2016) An extended U2AF(65)-RNA-binding domain recognizes the 3′ splice site signal. *Nat Commun* 7: 10950

39. Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* 40: D130–D135

40. Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcárcel J, Sattler M (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475: 408–411

41. Hastings ML, Allemand E, Duelli DM, Myers MP, Krainer AR (2007) Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). *PLoS One* 2: e538

42. Mai S, Qu X, Li P, Ma Q, Cao C, Liu X (2016) Global regulation of alternative RNA splicing by the SR-rich protein RBM39. *Biochem Biophys Acta* 1859: 1014–1024

43. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK *et al* (2018) The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46: D794–D801

44. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H *et al* (2014) Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol* 21: 997–1005

45. Královicová J, Ševcíková I, Stejskalová E, Obuca M, Hiller M, Stanek D, Vorechovský I (2018) PUF60-activated exons uncover altered 3′ splice-site selection by germline missense mutations in a single RRM. *Nucleic Acids Res* 46: 6166–6187

46. Singh R, Banerjee H, Green MR (2000) Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor sex-lethal. *RNA* 6: 901–911

47. Galganski L, Urbanek MO, Krzyzosiak WJ (2017) Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* 45: 10350–10368

48. Carmo-Fonseca M, Tollervey D, Pepperkok R, Barabino SM, Merdes A, Brunner C, Zamore PD, Green MR, Hurt E, Lamond AI (1991) Mammalian nuclei contain foci which are highly enriched in components of the pre-mRNA splicing machinery. *EMBO J* 10: 195–206

49. Gama-Carvalho M, Krauss RD, Chiang L, Valcárcel J, Green MR, Carmo-Fonseca M (1997) Targeting of U2AF65 to sites of active splicing in the nucleus. *J Cell Biol* 137: 975–987

50. Al-Ayoubi AM, Zheng H, Liu Y, Bai T, Eblen ST (2012) Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion. *Mol Cell Biol* 32: 2880–2893

51. Lallena MJJ, Chalmers KJ, Llamazares S, Lamond AI, Valcárcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3′ splice site recognition by SPF45. *Cell* 109: 285–296

52. Kralovicova J, Knut M, Cross NC, Vorechovsky I (2015) Identification of U2AF(35)-dependent exons by RNA-Seq reveals a link between 3′ splice-site organization and activity of U2AF-related proteins. *Nucleic Acids Res* 43: 3747–3763

53. Ying Y, Wang X-JJ, Vuong CK, Lin C-HH, Damianov A, Black DL (2017) Splicing activation by rbfox requires self-aggregation through its tyrosine-rich domain. *Cell* 170: 312–323

54. Gueroussov S, Weatheritt RJ, O'Hanlon D, Lin Z-YY, Narula A, Gingras A-CC, Blencowe BJ (2017) Regulatory expansion in mammals of multivalent hnRNP assemblies that globally control alternative splicing. *Cell* 170: 324–339

55. Zhang H, Elbaum-Garfinkle S, Langdon EM, Taylor N, Occhipinti P, Bridges AA, Brangwynne CP, Gladfelter AS (2015) RNA controls PolyQ protein phase transitions. *Mol Cell* 60: 220–230

56. Saha S, Weber CA, Nousch M, Adame-Arana O, Hoege C, Hein MY, Osborne-Nishimura E, Mahamid J, Jahnel M, Jawerth L *et al* (2016) Polar positioning of phase-separated liquid compartments in cells regulated by an mRNA competition mechanism. *Cell* 166: 1572–1584

57. Shepard PJ, Hertel KJ (2009) The SR protein family. *Genome Biol* 10: 242

58. Cáceres J, Misteli T, Screaton G, Spector D, Krainer A (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* 138: 225–238

59. Hedley M, Amrein H, Maniatis T (1995) An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci USA* 92: 11524–11528

60. Lee C-G, Zamore PD, Green MR, Hurwitz J (1993) RNA annealing activity is intrinsically associated with U2AF. *J Biol Chem* 268: 13472–13478

61. Yue B-G, Ajuh P, Akusjärvi G, Lamond AI, Kreivi J-P (2000) Functional coexpression of serine protein kinase SRPK1 and its substrate ASF/SF2 in *Escherichia coli*. *Nucleic Acids Res* 28: e14

62. Rudner DZ, Breger KS, Rio DC (1998) Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *Drosophila* subunit is dispensable *in vivo*. *Genes Dev* 12: 1010–1021

63. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155: 1049–1060

64. Ujvári A, Luse DS (2006) RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* 13: 49–54

65. Robert F, Blanchette M, Maes O, Chabot B, Coulombe B (2002) A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J Biol Chem* 277: 9302–9306

66. David CJ, Boyne AR, Millhouse SR, Manley JL (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev* 25: 972–983

67. Herzel L, Ottoz DSMS, Alpert T, Neugebauer KM (2017) Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* 18: 637–650

68. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BAA *et al* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46: W537–W544

69. Rodor J, Pan Q, Blencowe BJ, Eyras E, Cáceres JF (2016) The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation. *RNA* 22: 1411–1426

70. Clark F, Thanaraj TA (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum Mol Genet* 11: 451–464

71. Younis I, Berg M, Kaida D, Dittmar K, Wang C, Dreyfuss G (2010) Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol Cell Biol* 30: 1718–1728

72. Manceau V, Kremmer E, Nabel EG, Maucuer A (2012) The protein kinase KIS impacts gene expression during development and fear conditioning in adult mice. *PLoS One* 7: e43946